

Requested Patent: JP9282105A

Title:

DISK STORAGE SYSTEM AND RECOVERY METHOD FOR ERROR CORRECTION
DATA PREPARATION PROCESSING THEREFOR ;

Abstracted Patent: JP9282105 ;

Publication Date: 1997-10-31 ;

Inventor(s):

KITAMURA MANABU; YAMAMOTO AKIRA; YAMAMOTO YASUTOMO; SATO
TAKAO ;

Applicant(s): HITACHI LTD ;

Application Number: JP19960086086 19960409 ;

Priority Number(s): ;

IPC Classification: G06F3/06; G06F3/06; G06F3/06; G06F12/16 ;

Equivalents: ;

ABSTRACT:

PROBLEM TO BE SOLVED: To guarantee data in the case of occurrence of a fault during preparing a parity in a storage system for tentatively storing updating data sent from a processor in a cache memory, preparing a new parity and then storing the updating data and the parity in a storage device. SOLUTION: A nonvolatile cache 33 is provided with respective areas for storing the data after updating and before updating and the parities after updating and before updating. A parity preparation control part 38 prepares the parity after updating from the data before and after updating and the parity before updating, then turns ON a nonvolatile identifier for indicating parity preparation completion and copies the new parity and the data after updating to respective areas before updating. After the processing of a storage controller 3 is interrupted, a recovery control part 39 controls whether to re-execute a parity preparation processing from the start or re-execute it from the processing after turning ON the identifier by the OFF or ON of the identifier.

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平9-282105

(43)公開日 平成9年(1997)10月31日

(51)Int.Cl. ⁶	識別記号	序内整理番号	F I	技術表示箇所
G 06 F 3/06	305		G 06 F 3/06	305 C
	302			302 A
	540			540
12/16	320	7623-5B	12/16	320 L

審査請求 未請求 請求項の数9 O.L (全15頁)

(21)出願番号	特願平8-86086	(71)出願人	000005108 株式会社日立製作所 東京都千代田区神田駿河台四丁目6番地
(22)出願日	平成8年(1996)4月9日	(72)発明者	北村 学 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内
		(72)発明者	山本 彰 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内
		(72)発明者	山本 康友 神奈川県川崎市麻生区王禅寺1099番地 株式会社日立製作所システム開発研究所内
		(74)代理人	弁理士 薄田 利幸
			最終頁に続く

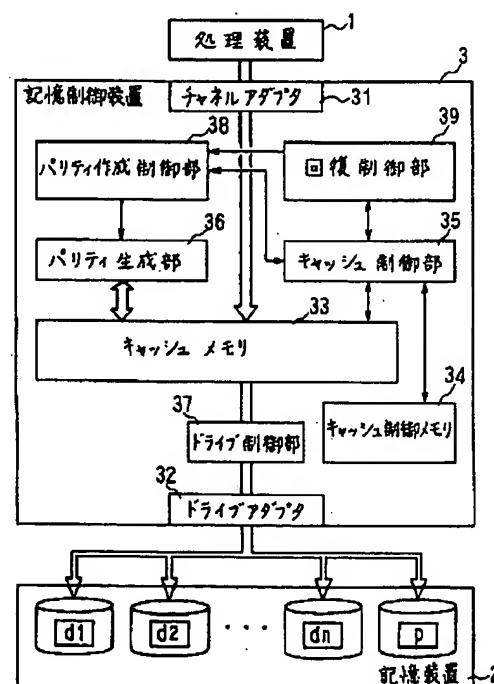
(54)【発明の名称】 ディスク記憶システム及びその誤り訂正データ作成処理の回復方法

(57)【要約】

【課題】 处理装置1から送られた更新データを一旦キッシュメモリ33に格納し、新しいパリティを作成してから更新データとパリティとを記憶装置2に格納する記憶システムにおいて、パリティ作成中に障害が発生した場合のデータの保証をする。

【解決手段】 不揮発のキッシュ33は、更新後及び更新前のデータ、更新後及び更新前のパリティを格納するそれぞれの領域を有する。パリティ作成制御部38は、更新前後のデータ及び更新前のパリティから更新後のパリティを作成した後、パリティ作成完了を示す不揮発の識別子をオンにして新しいパリティと更新後データをそれぞれの更新前領域にコピーする。記憶制御装置3の処理が中断した後、回復制御部39は識別子のオフ又はオンによってパリティ作成処理を始めから再実行するか又は識別子をオンにした後の処理から再実行するかを制御する。

図1



【特許請求の範囲】

【請求項1】データ及び誤り訂正データを格納するディスク記憶装置と、処理装置と該ディスク記憶装置との間に介在し、該処理装置と該ディスク記憶装置との間のデータ転送を制御する記憶制御装置とを有するディスク記憶システムであつて、該記憶制御装置は、該ディスク記憶装置に反映していない更新データを格納する第1の領域と、該更新データに対応し該ディスク記憶装置に反映済みのデータを格納する第2の領域と、該更新データに對応する誤り訂正データを格納する第3の領域と、該反映済みのデータに対応する誤り訂正データを格納する第4の領域とを有する不揮発性のキャッシュメモリと、第1の領域に格納される更新データ、第2の領域に格納される反映済みのデータ及び第4の領域の誤り訂正データから更新後の誤り訂正データを作成して第3の領域に格納する第1の処理ステップと、第3の領域の更新後の誤り訂正データを第4の領域へ上書きし、第1の領域の更新データを第2の領域へ上書きする第2の処理ステップとを実行する処理手段とを有するディスク記憶システムにおいて、

該記憶制御装置は、さらに第1の処理ステップが終了したか否かの状態を記憶する不揮発性の識別子と、第1の処理ステップが終了したとき該識別子を終了状態に設定する処理手段と、障害によって該記憶制御装置の処理が中断された後、該識別子を参照して終了状態でなければ該処理手段の第1の処理ステップを再実行し、終了状態であれば該処理手段の第2の処理ステップを再実行するよう制御する回復処理手段とを有するディスク記憶システム。

【請求項2】第3の領域及び第4の領域の誤り訂正データに対応して第1の領域及び第2の領域が複数個存在し、各第2の領域及び第4の領域の内容がそれぞれディスクアレイを構成する別々のディスク記憶装置に格納されることを特徴とする請求項1記載のディスク記憶システム。

【請求項3】処理装置とデータ及び誤り訂正データを格納するディスク記憶装置との間に介在し、該処理装置と該ディスク記憶装置との間のデータ転送を制御する記憶制御装置であつて、該記憶装置に反映していない更新データを格納する第1の領域と、該更新データに対応し該ディスク記憶装置に反映済みのデータを格納する第2の領域と、該更新データに対応する誤り訂正データを格納する第3の領域と、該反映済みのデータに対応する誤り訂正データを有する不揮発性のキャッシュメモリと、第1の領域に格納される更新データ、第2の領域に格納される反映済みのデータ及び第4の領域の誤り訂正データから更新後の誤り訂正データを作成して第3の領域に格納する第1の処理ステップと、第3の領域の更新後の誤り訂正データを第4の領域へ上書きし、第1の領域の更新データを第2の領域へ上書き

する第2の処理ステップとを実行する処理手段とを有する記憶制御装置において、

該記憶制御装置は、さらに第1の処理ステップが終了したか否かの状態を記憶する不揮発性の識別子と、第1の処理ステップが終了したとき該識別子を終了状態に設定する処理手段と、障害によって該記憶制御装置の処理が中断された後、該識別子を参照して終了状態でなければ該処理手段の第1の処理ステップを再実行し、終了状態であれば該処理手段の第2の処理ステップを再実行するよう制御する回復処理手段とを有する記憶制御装置。

【請求項4】該記憶装置は、複数のディスク記憶装置から構成されるディスクアレイであり、第3の領域及び第4の領域の誤り訂正データに対応して第1の領域及び第2の領域が複数個存在し、各第2の領域及び第4の領域の内容が別々のディスク記憶装置に格納されることを特徴とする請求項3記載の記憶制御装置。

【請求項5】該キャッシュメモリは、第1の領域、第2の領域、第3の領域及び第4の領域を有する第1のキャッシュメモリと、第1のキャッシュメモリと同じ内容の情報を格納する第2のキャッシュメモリとから構成され、該処理手段は、第1のキャッシュメモリと第2のキャッシュメモリのうち正常なキャッシュメモリについて第1の処理ステップ及び第2の処理ステップを実行し、該回復処理手段は、障害によって該記憶制御装置の処理が中断された後、第1のキャッシュメモリと第2のキャッシュメモリのうち正常なキャッシュメモリについて再実行するよう制御することを特徴とする請求項1記載のディスク記憶システム。

【請求項6】該キャッシュメモリは、第1の領域、第2の領域、第3の領域及び第4の領域を有する第1のキャッシュメモリと、第1のキャッシュメモリと同じ内容の情報を格納する第2のキャッシュメモリとから構成され、該処理手段は、第1のキャッシュメモリと第2のキャッシュメモリのうち正常なキャッシュメモリについて第1の処理ステップ及び第2の処理ステップを実行し、該回復処理手段は、障害によって該記憶制御装置の処理が中断された後、第1のキャッシュメモリと第2のキャッシュメモリのうち正常なキャッシュメモリについて再実行するよう制御することを特徴とする請求項3記載の記憶制御装置。

【請求項7】データ及び誤り訂正データを格納するディスク記憶装置と、処理装置と該ディスク記憶装置との間に介在し、該処理装置と該ディスク記憶装置との間のデータ転送を制御する記憶制御装置とを有するディスク記憶システムであつて、該記憶制御装置は、該ディスク記憶装置に反映していない更新データを格納する第1の領域と、該更新データに対応し該ディスク記憶装置に反映済みのデータを格納する第2の領域と、該更新データに對応する誤り訂正データを格納する第3の領域と、該反映済みのデータに対応する誤り訂正データを格納する第4の領域と

4の領域とを有する不揮発性のキャッシュメモリを含むディスク記憶システムの誤り訂正データ作成方法において、

第1の領域に格納される更新データ、第2の領域に格納される反映済みのデータ及び第4の領域の誤り訂正データから更新後の誤り訂正データを作成して第3の領域に格納する第1のステップと、

第1のステップが終了したときその状態を示す識別子を不揮発性の記憶装置に格納する第2のステップと、第3の領域の更新後の誤り訂正データを第4の領域へ上書きし、第1の領域の更新データを第2の領域へ上書きする第3のステップと、

障害によって該記憶制御装置の処理が中断された後、該識別子を参照して終了状態でなければ第1のステップを再実行し、終了状態であれば第3のステップを再実行する第4のステップとを有することを特徴とするディスク記憶システムの誤り訂正データ作成処理の回復方法。

【請求項8】第3の領域及び第4の領域の誤り訂正データに対応して第1の領域及び第2の領域が複数個存在し、各第2の領域及び第4の領域の内容がそれぞれディスクアレイを構成する別々のディスク記憶装置に格納されることを特徴とする請求項7記載のディスク記憶システムの誤り訂正データ作成処理の回復方法。

【請求項9】該キャッシュメモリは、第1の領域、第2の領域、第3の領域及び第4の領域を有する第1のキャッシュメモリと、第1のキャッシュメモリと同じ内容の情報を格納する第2のキャッシュメモリとから構成され、第1のステップから第4のステップまでの処理は、第1のキャッシュメモリと第2のキャッシュメモリのうち正常なキャッシュメモリについて行うことを特徴とする請求項7記載のディスク記憶システムの誤り訂正データ

$$P = d_1 (+) d_2 (+) \dots$$

ただし (+) は、排他的論理和を表す。排他的論理和は、各項を2進数に変換した時の各桁をそれぞれ参照し、1が奇数個の時は1、偶数個の時は0になる。例えばデータを3分割にして書き込む場合、 $d_1 = 01$, $d_2 = 11$, $d_3 = 00$ の時(それぞれ2進数)には P は10になる。この d_1 , d_2 , ..., d_n と P のデータの集合をパリティグループといい、 P は常に(1)式に

$$P = D (+) d_1 (+) P$$

特開平4-245352号公報には、記憶装置であるディスク装置がRAID構成をとりキャッシュメモリを有する記憶制御装置のライト処理の方法が記載されている。キャッシュメモリを有する記憶制御装置は、中央処理装置からのライト指示に対し記憶制御装置内のキャッシュメモリにライトデータを格納した時点で中央処理装置に対してライト処理終了を報告する。そして記憶制御装置は、中央処理装置からのライト指示とは非同期にディスク装置へライトデータを格納する。この制御をライトアフタ制御といい、ディスク装置へのデータ転送を行

タ作成処理の回復方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、ディスク記憶システム及びその誤り訂正データ作成処理に係わり、特に誤り訂正データ作成処理の回復方法に関する。

【0002】

【従来の技術】従来、デビット・A・パターソンらは、高速に大量のデータを複数のディスクにアクセスするディスク装置のディスク故障時におけるデータの冗長性を実現するディスクアレイ装置についての論文「低価格ディスクの冗長アレイ(RAID)の一実例(A Case for Redundant Arrays of Inexpensive Disks(RAID))」を1988年6月1日から3日にシカゴで開催されたACM SIGMODで発表している。提案されたディスクアレイ装置は、RAID (Redundant Arrays of Inexpensive Disks) と略称される。RAIDの仕組みを簡単に説明すると次のようなものである。

【0003】RAIDでは、ライトするデータを複数のデータに分割し、この分割した複数のデータについて排他的論理和をとることによってパリティを計算する。そして分割したデータとパリティを並列に複数のディスクに対して書き込む。ライトデータを n 個に分割した場合、それらのデータを d_1 , d_2 , ..., d_n とするときこの分割したデータからパリティ P を生成するには次の(1)式に従って行い、パリティ P の作成後に、 d_1 , d_2 , ..., d_n と P はそれぞれ別のディスクに書き込まれる。

【0004】

$$(+) d_n \quad (1)$$

従った値にならなければならぬ。そのため例えば既に書き込みの行われた領域 d_1 のみに別のデータ D が上書きされた場合でも、パリティ P について新規パリティ P' を算出して D とともに P もディスクに書き込む必要がある。その場合は次の(2)式によりパリティを算出して D と P をディスクに書き込む操作をする。

【0005】

$$(2)$$

う前に中央処理装置に対してライト処理終了を報告するため、中央処理装置への高速な応答が可能になる。

【0006】この記憶制御装置は、パリティを作成するのに必要十分なデータをキャッシュメモリ上に格納した後パリティを作成する。以下ライト処理について説明する。ディスクに書き込み済みのデータ d_1 の一部又は全部に対して書き込み要求が来てデータ d_1 に対する更新データ D がキャッシュメモリ上に書き込まれると、記憶制御装置は中央処理装置へライト完了報告を行い、その後に非同期にパリティの作成を行う。記憶制御装置は上

式(2)に従ってパリティを作成するため、旧データd1及び旧パリティpの置かれる領域と新たに作成されるパリティPを置く領域をキャッシュメモリ上に確保して、ディスクからd1及びpを読み込む。それから(2)式に従い新規にパリティPを作成して確保したキャッシュメモリ上の領域に格納する。その後、更新データDと新規作成パリティPをキャッシュメモリ上のd1及びpのある領域へ上書きする処理を行ってから、DとPをディスク装置へ書き込む。このときd1とpのデータが既にキャッシュメモリ上にある場合には、新たにキャッシュを確保する処理とd1とpをディスクから読み込む処理が省略できるために処理が高速化される利点がある。

【0007】

【発明が解決しようとする課題】RAID構成の記憶装置では、前述の通りデータの一部を書き換えるとともにそれに対応するパリティグループのパリティデータも常に前述の(1)式に従っていなければならぬので、データの更新と同時に書き換える必要がある。ただし実際にはデータの書き換えとパリティの書き換えは完全に同時に実行できるわけではなく、一時的にはパリティだけ又はデータだけが先にディスクに書き込まれているという状態が存在する。これは記憶制御装置がキャッシュメモリを有する場合にキャッシュメモリの内容についても同じことが言える。ライトデータDからパリティPを作成してd1とpの置かれているキャッシュメモリの領域にコピーする時、Dだけがコピーされた時点で停電等による処理中断が発生した場合、パリティ作成前にはd1とpが置かれていた領域はDとpが存在するという状態になる。そのために例えばこの状態から同じパリティグループ内の別のデータd2に対する更新データD2からのパリティP2の作成を行うと、本来は

$$P2 = d2 (+) D2 (+) P$$

を行わなければならないのに、Pの更新前パリティデータpを使った

$$d2 (+) D2 (+) P$$

を行って間違ったパリティを作成してしまうため、このようなパリティ作成途中での停電等による処理中断や障害発生でのデータ損失の問題は十分考慮される必要がある。しかしながら従来技術では、パリティ作成中の停電等による処理中断が発生した場合、特に更新データDとDに対するパリティデータPを、更新前データd1とd1に対するパリティpのおかれている領域にコピーしている際のデータ損失やその際の回復処理について考慮されていない。

【0008】このd1やpがディスク装置に書き込み済みの状態の場合、d1及びpの置かれているキャッシュメモリにディスク装置からその内容を再度読み込んで、パリティの作成処理を始めからやり直しをすれば良いことは容易に考えられるので考慮する必要はない。しかし

ライトアフタ制御ではキャッシュメモリにライトデータを格納した時点で中央処理装置に対してライト処理終了を報告するため、ライトデータDからパリティPを作成して、それらがd1やpのある領域にコピーされた直後に同じ領域にさらに新しいライトデータD'がキャッシュメモリ上に書き込まれてそれに対する新規パリティP'を作成する処理が始まるこもあり得る。その時に上の例のようにD'をDのある領域にコピーした直後で障害が発生した場合にパリティ作成をやり直そうとしても、DとPはディスク上には置かれていでないデータなので読み込むことはできず、やり直しはできない。

【0009】本発明の目的は、このようなディスク装置への書き込みが行われていないデータに対する更新データからパリティ作成を行っている途中に障害が発生した場合のデータ保証をすることにある。

【0010】

【課題を解決するための手段】本発明は、不揮発のキャッシュメモリ上に置かれた更新データD、ディスク記憶装置に反映済みの対応するデータd及びdに対応する誤り訂正データpから更新後の誤り訂正データPを作成する第1のステップと、第1のステップが終了したときその状態を示す識別子を不揮発性の記憶装置に格納する第2のステップと、誤り訂正データPをpの領域へ上書きし、更新データDをdの領域へ上書きする第3のステップと、障害によって記憶制御装置の処理が中断された後、この識別子を参照して終了状態でなければ第1のステップを再実行し、終了状態であれば第3のステップを再実行する第4のステップとを有することを特徴とするディスク記憶システムの誤り訂正データ作成処理の回復方法を特徴とする。

【0011】本発明によれば、誤り訂正データの作成が終了したか否かを境界にして回復処理が再実行を開始するポイントを変えており、いずれの場合にもデータ及びパリティの正しいことが保証され、再開始後の処理においてもディスク記憶システムは間違ったパリティを作成することはない。

【0012】なお同一パリティグループ内に更新データがD1, D2, ... のように複数ある場合には、第1のステップでは更新データD1, D2, ... 、それぞれ反映済みの対応するデータd1, d2, ... 及び誤り訂正データpから更新後の誤り訂正データPを作成し、第3のステップでは更新データD1, D2, ... をそれぞれd1, d2, ... へ上書きし、その他の処理は上記の通りである。

【0013】

【発明の実施の形態】以下、本発明の実施形態について図面を用いて説明する。

【0014】(1) 第1の実施形態

図1は、第1の実施形態のディスク記憶システムの構成図である。ディスク記憶システムは、記憶制御装置3及

び記憶装置2から構成される。処理装置1は、記憶装置2に対してリード／ライト要求を発行する情報処理装置である。記憶装置2は、複数のディスク記憶装置を有し、ディスクアレイを構成する。記憶制御装置3は、処理装置1及び記憶装置2に接続され、チャネルアダプタ31、ドライブアダプタ32、キャッシュメモリ33、キャッシュ制御メモリ34、キャッシュ制御部35、パリティ生成部36、ドライブ制御部37、パリティ作成制御部38及び回復制御部39から構成される。図1で太い矢印はデータの流れを示し、細い矢印は制御情報の流れを示す。ただし図面を簡単にするために本発明と関連の少ない制御の流れについて省略している。チャネルアダプタ31は、チャネルインターフェースによって処理装置1のチャネルに接続され、処理装置1と情報の送受信を行う。ドライブアダプタ32は、ドライブインターフェースによって記憶装置2に接続され、記憶装置2と情報の送受信を行う。キャッシュメモリ33は、バッテリにより不揮発化された記憶装置であり、記憶装置2に書き込むデータ及びその誤り訂正データを一時的に格納する。以下誤り訂正データをパリティと呼ぶ。キャッシュ制御メモリ34は、不揮発化された記憶装置であり、キャッシュメモリ33上のデータ及びパリティを管理する情報を格納する記憶装置である。キャッシュ制御部35は、キャッシュメモリ33上の記憶領域を管理し、キャッシュ制御メモリ34上に管理情報を作成し、更新する。パリティ生成部36は、データ更新に伴ってキャッシュメモリ33上のデータ及びパリティから新しいパリティを生成する。ドライブ制御部37は、記憶装置2内のディスク記憶装置とキャッシュメモリ33との間のデータの転送を制御する。パリティ作成制御部38は、パリティ作成処理を制御し、そのステータスを管理する。回復制御部39は、停電等の障害が発生してパリティの作成が中断したとき、パリティ作成処理を回復する。

【0015】ここで以下の説明に使用する用語の説明をする。レコードとは、処理装置1と記憶制御装置3との間で転送するデータの単位である。スロットとは、キャッシュメモリ33上の記憶領域の単位であるとともに、処理装置1から来たライトデータを記憶装置2の各ディスク記憶装置に分割する際の大きさの単位である。スロットのサイズはレコードのサイズより大きく、スロットは複数のレコードを収容可能である。すべてのスロットは同じ大きさである。各スロットにはユニークなスロット番号が付与される。処理装置1からリード／ライトされるデータが置かれるスロットをデータスロットといい、このデータスロットから作成されるパリティが置かれるスロットをパリティスロットという。スロットはさらに細かいブロックに分割される。ブロックはディスク記憶装置と記憶制御装置3との間でデータを転送するときの最小単位である。同一スロット内の各ブロックにはブロック番号が付与される。スロット番号とブロック番

号が与えられると、特定のディスク記憶装置上のデータ格納場所が決定される。1つのパリティを作成するときに対象となる複数スロットに亘るデータの集合をそのパリティとともにまとめてパリティグループという。キャッシュメモリ33上のデータのうち記憶装置2上のデータと一致しているデータをクリーンデータという。記憶装置2に反映されておらず、キャッシュメモリ33上にのみ記憶されている更新データをダーティデータという。ダーティデータのうちパリティの作成されているデータを物理ダーティデータ、パリティの作成されていないデータをホストダーティデータという。

【0016】処理装置1から1レコード又は複数レコードから成るデータのライト要求が来たとき、記憶制御装置3はこのデータの大きさが1スロットに収まる範囲の大きさであればそのまま、複数スロットにまたがる大きさの場合にはスロット境界でデータd1, d2, ..., dnに分割し、一旦キャッシュメモリ33に記憶する。分割したデータはそれぞれキャッシュメモリ33上の連続するスロット番号をもつ別々のスロットに格納される。1つのレコードはスロット内の1つ又は複数の連続するブロックを占有する。キャッシュメモリ33へのデータ書き込みが終了した時点で記憶制御装置3は、処理装置1にライト終了を報告し、その後処理装置1のライト指示とは非同期に单一のデータスロット又は同一のパリティグループに属する複数のデータスロットからパリティを作成する。パリティ作成制御部38はパリティ生成部36にパリティの生成を要求する。パリティ生成部36は、キャッシュメモリ33上のデータを読み出してパリティを作成し、キャッシュメモリ33上のパリティスロットにこのパリティを書き込む。パリティの作成が終了した後、ドライブ制御部37は各データスロット内のデータとパリティスロット内のパリティを各々別々のディスク記憶装置に格納する。

【0017】図2は、キャッシュメモリ33内に確保されるスロットの構成とキャッシュ制御メモリ34内に確保されるスロット管理情報の構成を示す図である。各スロット42は、それぞれ分割したデータ及びパリティを含む。1つのスロット42は、ライト面44とリード面45から構成される。ライト面44は処理装置1からのライト要求によって受け取ったデータを格納する領域、リード面45はディスク記憶装置に書き込むデータを格納する領域である。ライト面44はホストダーティデータを格納し、リード面45は物理ダーティデータ又はクリーンデータを格納する。ライト面44又はリード面45は、さらに複数のブロック46に分割されている。分割したデータ及びパリティは、それぞれライト面44及びリード面45内の1つ又は複数のブロックを占有する。

【0018】キャッシュ制御メモリ34内に確保されるスロット管理情報43は、スロット番号、ライト面ポイ

ンタ、リード面ポインタ及びスロットステータス47から構成され、それぞれ対応するスロットに置かれる分割されたデータ又はパリティを管理する。スロット番号は管理するスロットの番号である。ライト面ポインタは、ライト面44内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納する。またリード面ポインタは、リード面45内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納する。スロットステータス47は、スロット内で管理しているデータの状態を示す情報を保有しており、チャネルライト中、ディスクリード中、ディスクライト中、パリティ作成中、パリティ作成完了、ホストダーティ及び物理ダーティを示す識別子から構成される。各識別子は、その状態がオンかオフかの2値をもつ。チャネルライト中、ディスクリード中、ディスクライト中は、それぞれ処理装置1からキャッシュメモリ33へのデータ転送中、記憶装置2からキャッシュメモリ33へのデータ転送中、キャッシュメモリ33から記憶装置2へのデータ転送中の状態を示す。

【0019】処理装置1からライト要求が来ると、記憶制御装置3は指定されたシリンド番号／トラック番号／レコード番号のようなディスク・アドレスをスロット番号／ブロック番号に変換し、処理装置1から受け取った1つ又は複数のレコードをそのまま又はn個のデータに分割する。次にキャッシュ制御部35はキャッシュ制御メモリ34上にこのスロット番号から始まるn個のスロット管理情報43を作成する。この操作をスロット確保という。次にキャッシュ制御部35は、キャッシュメモリ33の空いている領域からn個のスロット42のライト面44を確保し、各ライト面44上でレコードが開始するブロックのアドレスとブロック数をスロット管理情報43のライト面ポインタに格納する。この操作をキャッシュ確保という。キャッシュ確保をした後に記憶制御装置3は、各スロット42上のブロックに分割したデータの各々を書き込む。次にパリティ作成制御部38は、キャッシュ制御部35を介してパリティスロットとライト面、リード面のキャッシュを確保し、パリティ生成部36はパリティを作成又は更新してパリティスロットのライト面に書き込む。次にパリティ作成制御部38はパリティスロットのライト面44のパリティをリード面45にコピーし、各データスロットのライト面44のデータをリード面45にコピーする。この後スロット管理情報43のライト面ポインタの内容をクリアして確保したキャッシュをポインタから外す。この操作をキャッシュ解放という。次にデータスロットのリード面45上のデータ及びこのパリティグループに属するパリティスロットのリード面45上のパリティが記憶装置2に書き込まれ、スロット管理情報43のリード面ポインタの内容をクリアして各スロットのリード面のキャッシュを解放する。一方処理装置1からリード要求が来ると、スロット

42のリード面のキャッシュ確保をして記憶装置2からデータ及びパリティを読み込み、リードデータが不要になるとそのリード面のキャッシュを解放する。

【0020】次にパリティ作成制御部38及びパリティ生成部36が行うパリティ作成処理について説明する。パリティ作成処理は、大別するとフェーズ1とフェーズ2の2つの過程に分けられる。フェーズ1は、パリティ作成の対象となるデータスロットに対応してキャッシュメモリ33上にパリティスロットを確保し、データスロットの内容からパリティを作成する。フェーズ2は、データスロットのライト面のホストダーティデータとパリティスロットのライト面に作成したパリティを各々のスロットのリード面にコピーして物理ダーティデータとし、各スロットのライト面のキャッシュを解放する処理を行う。なお以下の説明を簡単にするために、1つのデータスロット内のレコードを更新する場合を主体として説明する。

【0021】図3及び図4は、フェーズ1の処理の流れを示すフローチャートである。パリティ作成制御部38は、キャッシュ制御部35を介して対象となるパリティグループについてパリティスロットのスロット管理情報43を作成してパリティスロットを確保し、そのスロット管理情報43のスロットステータス47のうち「パリティ作成中」と「ホストダーティ」をオンにする（ステップ101）。「パリティ作成中」は、スロットが他の処理によって操作されることを禁止するために用い、「ホストダーティ」状態は回復処理を行う時にこの状態を参照して処理を決定するために用いられる情報である。次にそのパリティスロットのリード面とライト面のキャッシュメモリが確保されているか否かをチェックし、確保されていない場合にはキャッシュ制御部35を介して確保する（ステップ102）。次に同一パリティグループ中のデータスロットについて、そのデータスロットに対応するスロット管理情報43のスロットステータス47が「ホストダーティ」であるデータスロットを選択し、選択したデータスロットに対応するスロットステータス47のうちの「パリティ作成中」をオンにし（ステップ103）、そのデータスロットのリード面とライト面のキャッシュメモリが確保されているか否かをチェックし、確保されていない場合にはキャッシュ制御部35を介して確保する（ステップ104）。次にパリティ作成対象のデータスロットに対応するスロット管理情報43のライト面ポインタを参照し、ホストダーティデータの存在する範囲を算出する（ステップ105）。すなわちデータスロットのライト面でレコードが開始するブロックのアドレスとブロック数を得る。次にステップ105で算出した範囲に対応するリード面に物理ダーティ又はクリーンデータがあるかどうかを判定し、存在しない場合には（ステップ106NO）、記憶装置2からクリーンデータをリード面に読み込む（ステップ10

7)。次にステップ105で算出したデータスロットのライト面のデータの存在する範囲について、対応するパリティスロットのリード面にパリティが存在するか確認する(ステップ108)。存在しない場合には(ステップ108No)、記憶装置2からその範囲のクリーンデータをパリティスロットのリード面に読み込む(ステップ109)。次にパリティ作成制御部38はパリティ生成部36に要求してパリティを作成する(ステップ110)。パリティ生成部36は、データスロットのリード面、データスロットのライト面及びパリティスロットのリード面のデータを読み出して指定された範囲のパリティを作成し、新しいパリティをパリティスロットのライト面に書き込む。

【0022】図5及び図6は、フェーズ2の処理の流れを示すフローチャートである。ステップ110の処理が終了するとすぐにパリティスロットのスロット管理情報43のスロットステータス47の「パリティ作成完了」をオンにする(ステップ201)。スロット42のライト面へのデータ出力が終了した直後にパリティスロットのライト面に作成された新しいパリティをそのリード面の対応する一連のブロックへ上書き(コピー)する(ステップ202)。次にデータスロットのホストダーティデータをそのリード面の対応する一連のブロックへ上書き(コピー)する(ステップ203)。次にパリティスロットについて不要となったライト面の一連のブロックをスロットから解放し(ステップ204)、パリティスロットのスロット管理情報43のスロットステータス47の「ホストダーティ」をオフ、代わりに「物理ダーティ」をオンする(ステップ205)。次にデータスロットについて不要となったライト面の一連のブロックをスロットから解放し(ステップ206)、スロットステータス47の「ホストダーティ」をオフにし、「物理ダーティ」をオンにする(ステップ207)。最後にパリティスロットとデータスロットのスロットステータス47の「パリティ作成中」をオフにして(ステップ208)、パリティ作成処理を終了する。

【0023】なおパリティグループ内の1つのスロット内のレコードを更新する場合に、キャッシング上の同一パリティグループ内の他のデータスロットに「ホストダーティ」のまま残っているレコードがあれば、ステップ103～104の処理をすべてのデータスロットについて繰り返し、またステップ105～107の処理をすべてのデータスロットについて繰り返し、ステップ108～110はデータスロット内すべてのホストダーティのデータが存在する範囲についてパリティを作成する必要がある。またステップ203、206及び207もパリティ作成の対象となるすべてのデータスロットについて処理を行う必要がある。

【0024】次にパリティ作成処理中の障害発生により処理が中断された後のスロット回復処理について説明す

る。記憶制御装置3に障害が発生すると、すべての処理は中断されて不揮発化されているキャッシングメモリ33とキャッシング制御メモリ34以外は初期状態に戻る。記憶制御装置3の動作が初期状態から開始されると、まず回復処理が起動する。

【0025】図7は、回復制御部39の処理の流れを示すフローチャートである。回復制御部39は、スロットステータス47が「パリティ作成中」状態のパリティスロットと、そのパリティグループ内のデータスロットでスロットステータス47の「パリティ作成中」がオンになっているすべてのスロットを取り出す(ステップ1001)。次にパリティスロットのスロットステータス47の「パリティ作成完了」がオンでなければ(ステップ1002No)、対象とするスロットステータス47をパリティ作成前の状態に戻す(ステップ1003)。すなわち「パリティ作成中」がオン状態のデータスロットについてはこれをオフにし、パリティスロットについては「パリティ作成中」をオフにし、「ホストダーティ」もオフにすることによってパリティ作成前の状態に戻る。そしてパリティ作成処理をフェーズ1の最初から再実行すればよい(ステップ1004)。すなわちパリティ作成処理がフェーズ1の途中で終っていた場合、パリティ作成制御部38はリード面のデータに操作を行っていないので、再びパリティ作成処理を始めから行ってもデータの整合性が保たれており問題ない。一方パリティスロットのスロットステータス47の「パリティ作成完了」がオンであれば(ステップ1002Yes)、パリティスロットのスロットステータス47の「ホストダーティ」がオンであれば(ステップ1005Yes)、B以降(ステップ202以降)を実行する(ステップ1006)。パリティスロットのスロットステータス47の「ホストダーティ」がオフであれば(ステップ1005No)、C以降(ステップ205以降)を実行する(ステップ1007)。パリティスロットのスロットステータス47の「パリティ作成完了」がオンであればフェーズ1は完了しているから、フェーズ2だけを実行してスロットをパリティ作成完了の状態にすればデータの整合性は保たれる。ここでパリティスロットのスロットステータス47の「ホストダーティ」がオフであれば、ステップ204までの処理は済んでいるのであるから、C以降を再実行すればよい。またパリティスロットのスロットステータス47の「ホストダーティ」がオンであれば、ステップ205の処理まで達していないので、B以降を再実行する必要がある。なおパリティスロットのスロットステータス47の「パリティ作成完了」がオンであるときには、ステップ202のパリティのリード面へのコピーが終了している可能性があるから、フェーズ1の最初から再実行すると、以降のパリティ作成に際して間違ったパリティ作成を行う可能性があることが理解される。

【0026】(2) 第2の実施形態

以下、第1の実施形態をベースとしてキャッシュメモリが二重化されている場合のパリティ作成制御部38及び回復制御部39の処理について述べる。キャッシュが二重化されている場合には、パリティの作成中に一方のキャッシュメモリに障害が発生して使用できないとき他方のキャッシュメモリの内容を使って回復を行い、処理続行することが可能である。以下第1の実施形態に対応する構成要素がある場合には同じ参考番号を用いる。

【0027】図8は、第2の実施形態のディスク記憶システムの構成図である。キャッシュメモリ33-1及びキャッシュメモリ33-2は、二重化されたキャッシュメモリである。以下キャッシュメモリ33-1及びキャッシュメモリ33-2をそれぞれキャッシュメモリ1及びキャッシュメモリ2又はキャッシュ1及びキャッシュ2と略称することがある。キャッシュ制御部35に含まれるキャッシュ障害判定部351は、キャッシュメモリ33-1及びキャッシュメモリ33-2が正常か否かを判定し、片方のキャッシュに障害があって使用できないときには、正常な方のキャッシュを選択する。

【0028】図9は、キャッシュ制御メモリ34内に確保されるスロット管理情報の構成及びキャッシュメモリ33内のスロットとの関連を示す図である。スロット管理情報43は、キャッシュメモリ1上のスロットとキャッシュメモリ2上のスロットの両方を管理する。すなわちキャッシュ1のライト面ポインタはキャッシュ1上のライト面44内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納し、リード面ポインタはキャッシュ1上のリード面45内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納する。またキャッシュ2のライト面ポインタはキャッシュ2上のライト面44内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納し、リード面ポインタはキャッシュ2上のリード面45内のレコードが占有する最初のブロックのアドレスとそのブロック数を格納する。スロット番号及びスロットステータス47については第1の実施形態で説明した通りである。

【0029】処理装置1からライト要求が来ると、キャッシュ制御部35はキャッシュメモリ33-1及びキャッシュメモリ33-2のそれれについて1個又はn個のスロット42のライト面44を確保し、各ライト面44上のブロック・アドレスとブロック数をそのスロットに対応するスロット管理情報43のキャッシュ1のライト面ポインタ及びキャッシュ2のライト面ポインタに格納してキャッシュを確保する。その後、記憶制御装置3は同一のデータをキャッシュ1のライト面とキャッシュ2のライト面とに重複して書き込む。

【0030】図10及び図11は、キャッシュが二重化されており、キャッシュ1とキャッシュ2がともに正常であるときのパリティ作成制御部38及びパリティ生成

部36が行うパリティ作成処理のフェーズ1の処理の流れを示すフローチャートである。第1の実施形態と異なる点は、キャッシュを確保するときに、キャッシュ1、キャッシュ2の順に確保し(ステップ402及び404)、キャッシュにデータを読み込むときにもキャッシュ1、キャッシュ2の両方に読み込みする(ステップ407及び409)点である。またパリティ生成部36はキャッシュ1とキャッシュ2の両方についてそれぞれパリティを作成する(ステップ410)。

【0031】図12及び図13は、図10及び図11に続くフェーズ2の処理の流れを示すフローチャートである。第1の実施形態と異なる点は、ライト面のデータをリード面にコピーするとき、キャッシュ1とキャッシュ2の両方について行い(ステップ502及び503)、キャッシュのライト面を解放するとき、キャッシュ1とキャッシュ2の両方について行う(ステップ504及び506)点である。

【0032】図14は、第2の実施形態の回復制御部39の処理の概略を示すフローチャートである。回復制御部39は、キャッシュが障害状態にあるかどうかチェックする(ステップ2001)。両方とも正常ならば(ステップ2001 Yes)、図15に示すキャッシュが二重化されているときの回復処理を行う。キャッシュ1のみが正常ならば(ステップ2002 Yes)、キャッシュ1について回復処理を行う。キャッシュ2のみが正常ならば(ステップ2002 No)、キャッシュ2について回復処理を行う。キャッシュ1又はキャッシュ2についての回復処理は、それぞれのキャッシュについて第1の実施形態の回復処理(図7)と同じである。

【0033】図15は、キャッシュが二重化されているときの回復制御部39の回復処理の流れを示すフローチャートである。回復処理は基本的に第1の実施形態と同じであるが、キャッシュ1とキャッシュ2の両方について回復処理を行う。

【0034】第2の実施形態によれば、キャッシュ1、キャッシュ2のうち1つに異常が発生していて使用できないときには、キャッシュ障害判定部351は、正常な方のキャッシュを選択するので、パリティ作成制御部38は選択された正常のキャッシュについてパリティ作成処理を続行できる。記憶制御装置3に障害が発生して処理が中断されたとき、回復制御部39はこの正常な方のキャッシュについて回復処理を行う。

【0035】

【発明の効果】本発明によれば、誤り訂正データの作成が終了したか否かを境界にしてそれぞれ適切なポイントからパリティ作成処理を再実行するので、正しいデータ及びパリティが保証される。

【図面の簡単な説明】

【図1】第1の実施形態のディスク記憶システムの構成図である。

【図2】第1の実施形態のキャッシュメモリ33及びキャッシュ制御メモリ34の構成を示す図である。

【図3】第1の実施形態のパリティ作成処理(フェーズ1)の処理の流れを示すフローチャートである。

【図4】第1の実施形態のパリティ作成処理(フェーズ1)の処理の流れを示すフローチャート(続き)である。

【図5】第1の実施形態のパリティ作成処理(フェーズ2)の処理の流れを示すフローチャートである。

【図6】第1の実施形態のパリティ作成処理(フェーズ2)の処理の流れを示すフローチャート(続き)である。

【図7】第1の実施形態の回復制御部39の処理の流れを示すフローチャートである。

【図8】第2の実施形態のディスク記憶システムの構成図である。

【図9】第2の実施形態のキャッシュメモリ33及びキャッシュ制御メモリ34の構成を示す図である。

【図10】第2の実施形態のパリティ作成処理(フェ

ズ1)の処理の流れを示すフローチャートである。

【図11】第2の実施形態のパリティ作成処理(フェーズ1)の処理の流れを示すフローチャート(続き)である。

【図12】第2の実施形態のパリティ作成処理(フェーズ2)の処理の流れを示すフローチャートである。

【図13】第2の実施形態のパリティ作成処理(フェーズ2)の処理の流れを示すフローチャート(続き)である。

【図14】第2の実施形態の回復制御部39の処理の流れを示すフローチャートである。

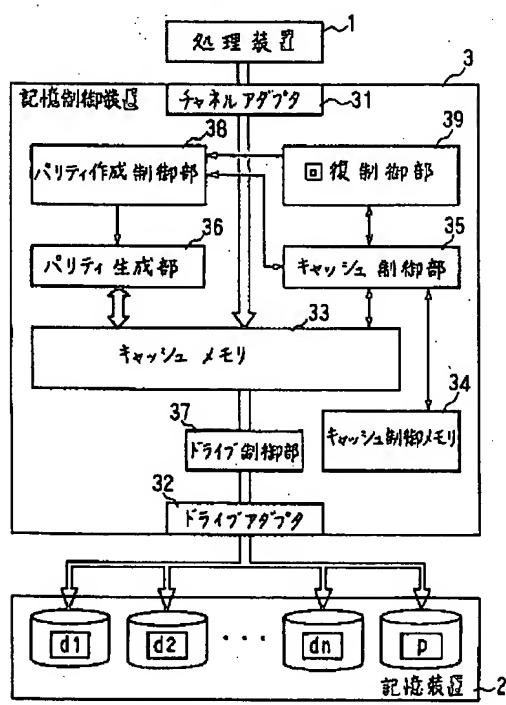
【図15】第2の実施形態の回復制御部39の処理の流れを示すフローチャート(続き)である。

【符号の説明】

3: 記憶制御装置、33: キャッシュメモリ、34: キャッシュ制御メモリ、36: パリティ生成部、38: パリティ作成制御部、39: 回復制御部、42: スロット、43: スロット管理情報、44: ライト面、45: リード面、47: スロットステータス

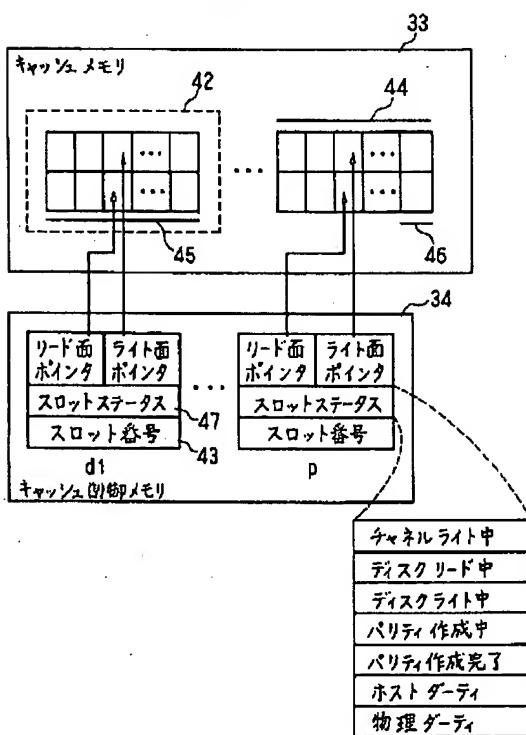
【図1】

図1



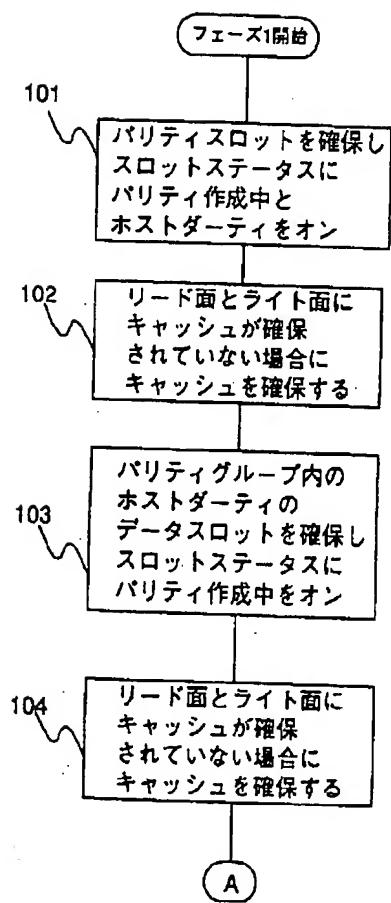
【図2】

図2



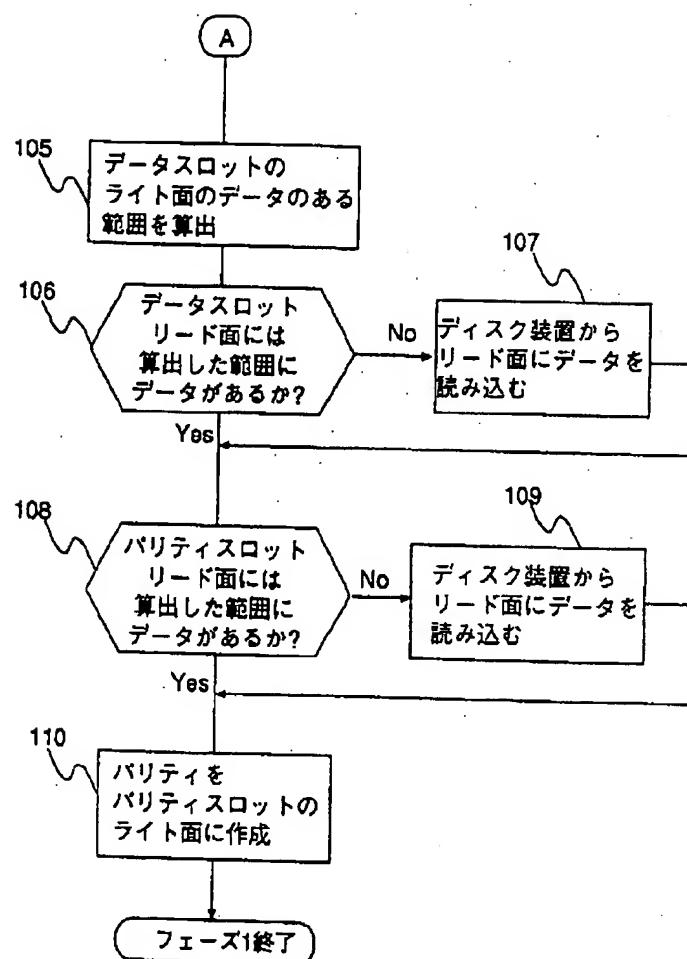
【図3】

図3



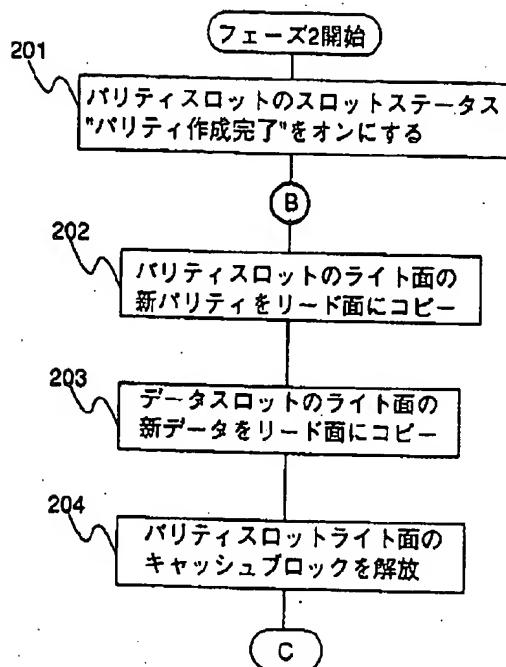
【図4】

図4



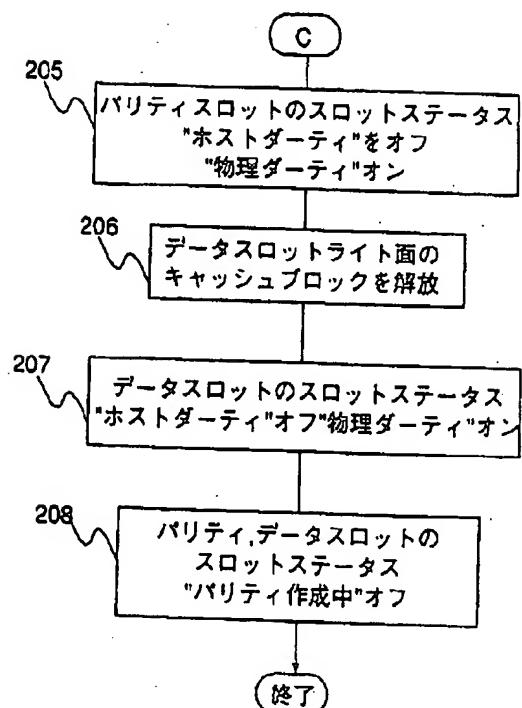
【図5】

図5



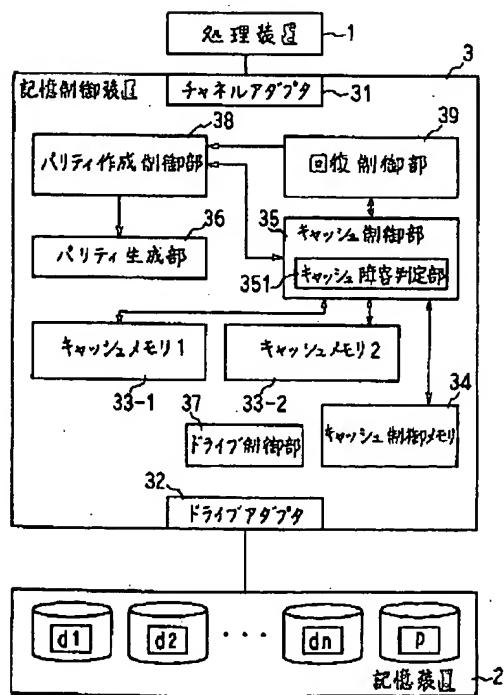
【図6】

図6



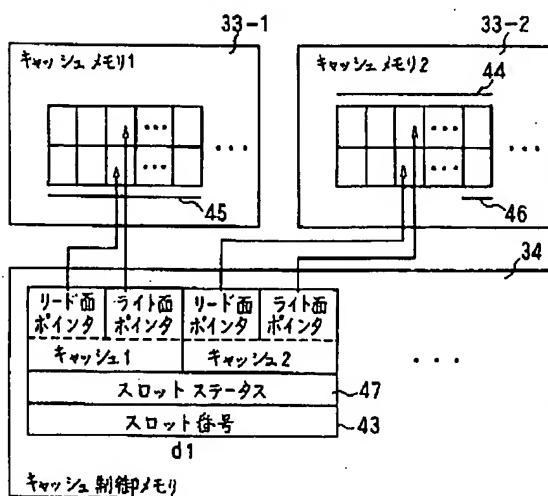
【図8】

図8



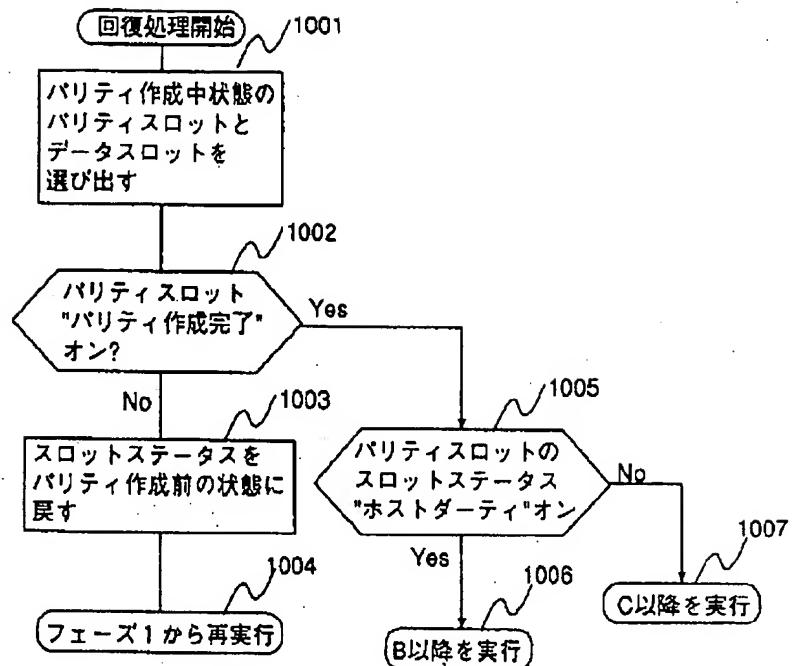
【図9】

図9



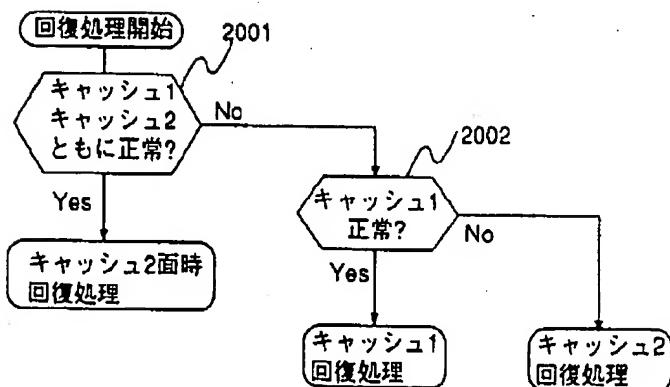
【図7】

図7



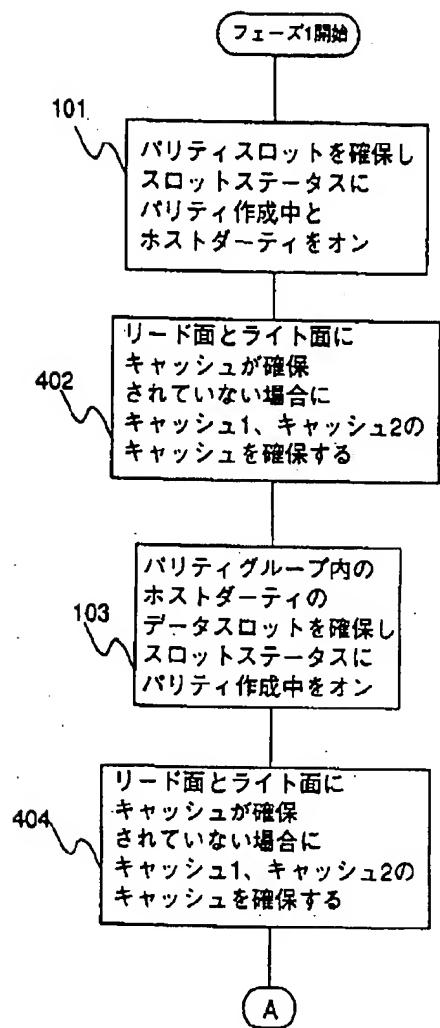
【図14】

図14



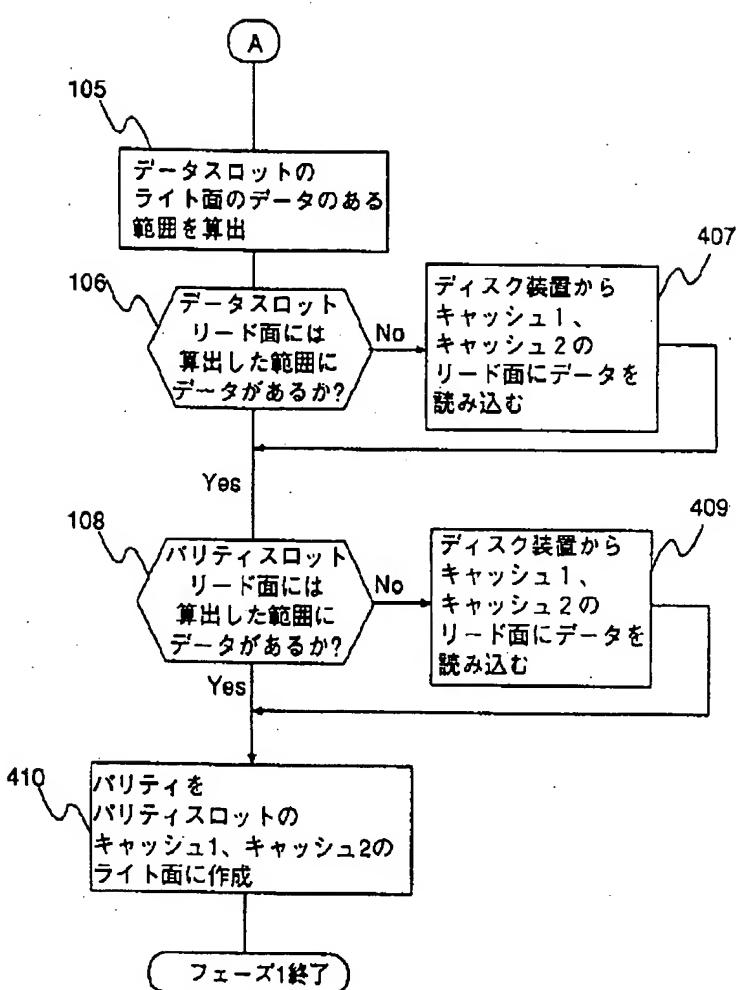
【図10】

図10



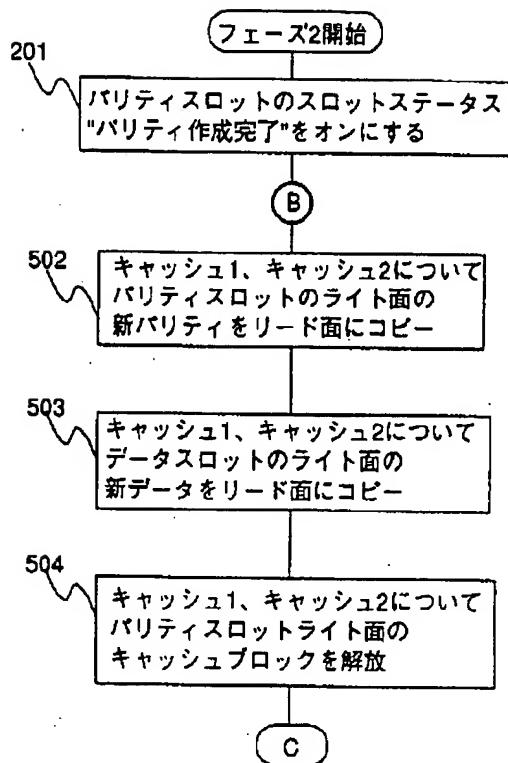
【図11】

図11



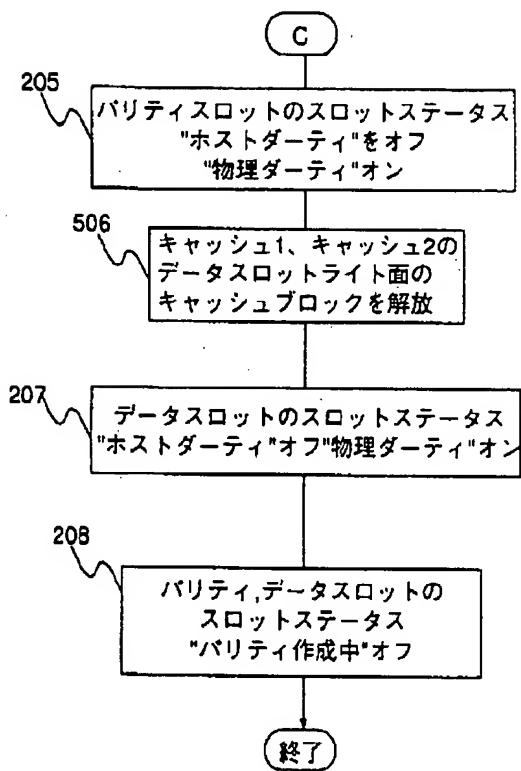
【図12】

図12



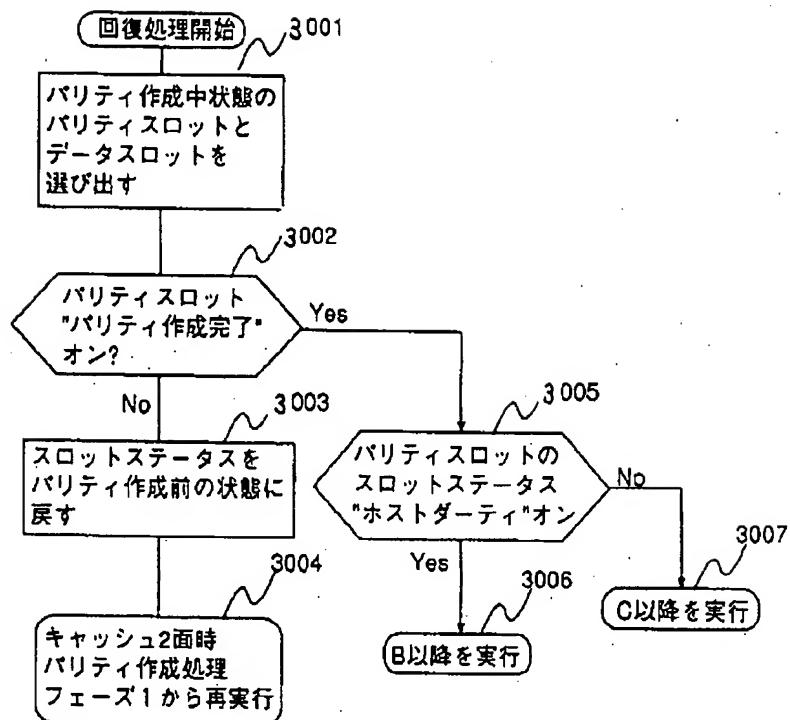
【図13】

図13



【図15】

図15



フロントページの続き

(72)発明者 佐藤 孝夫
 神奈川県小田原市国府津2880番地 株式会
 社日立製作所ストレージシステム事業部内